

Customer Segmentation Analysis - Exploratory Data Analysis Report

February 9, 2024

1 Project: Customer Segmentation and Analysis

1.1 Stage Two: Exploratory Data Analysis (EDA)

This is the second stage of the project, I delve into the cleaned sales data from the data wrangling stage of the project to understand the key aspects of the data. I use interactive visualizations to communicate key insights and trends, as well as summary statistics to highlight key aspects of the dataset, helping stakeholders understand the nature of the data.

In this analysis, I look at highlighting some Key Performance Indicators (KPIs) by answering questions on the data. These key questions are as follows:

1. **What proportion of transactions were captured as either sale or canceled?**
2. **What was the trend in transactions made over the recorded period?** For this question, trend in transactions will be grouped according to transaction types. That is **sale** or **canceled** and filtered by the year in which these transactions took place.
3. **From which countries were the most or least transactions recorded with respect to transaction type (sale or canceled)?** For this, I will look at the number of transactions recorded per country and the total amounts realised from these transactions with respect to the transaction type (sale or canceled). I will look at the countries that occupy the top and bottom 10 spots in the result of this analysis.
4. **During which time of day (morning, afternoon, evening) were transactions made?** For this interesting analysis, I will look at both transaction types (sale and canceled), the total amounts realised from each type during each time of day for the months in the years recorded. This will be filtered by the weeks in the months and will give stakeholders a deeper insight into the different types of transactions recorded and how these transactions fared during the morning, afternoon and evening hours of the day for each month and year.
5. **From which country did the online retail store have the most or least customers?**

For each of the above questions, key insights derived can be found below each chart used in visualizing the data explaining the findings.

- **NB:** It must be noted that the currency value of all transaction amounts is in pound sterling.

	InvoiceNo	StockCode	Description	Quantity	\
0	536365	85123	WHITE HANGING HEART T-LIGHT HOLDER	6	

1	536365	71053	WHITE METAL LANTERN	6
2	536365	84406	CREAM CUPID HEARTS COAT HANGER	8
3	536365	84029	KNITTED UNION FLAG HOT WATER BOTTLE	6
4	536365	84029	RED WOOLLY HOTTIE WHITE HEART.	6
5	536365	22752	SET 7 BABUSHKA NESTING BOXES	2
6	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6
7	536366	22633	HAND WARMER UNION JACK	6
8	536366	22632	HAND WARMER RED POLKA DOT	6
9	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32

	InvoiceDate	UnitPrice	CustomerID	Country	TransactionType	\
0	2010-12-01 08:26:00	2.55	17850	United Kingdom	sale	
1	2010-12-01 08:26:00	3.39	17850	United Kingdom	sale	
2	2010-12-01 08:26:00	2.75	17850	United Kingdom	sale	
3	2010-12-01 08:26:00	3.39	17850	United Kingdom	sale	
4	2010-12-01 08:26:00	3.39	17850	United Kingdom	sale	
5	2010-12-01 08:26:00	7.65	17850	United Kingdom	sale	
6	2010-12-01 08:26:00	4.25	17850	United Kingdom	sale	
7	2010-12-01 08:28:00	1.85	17850	United Kingdom	sale	
8	2010-12-01 08:28:00	1.85	17850	United Kingdom	sale	
9	2010-12-01 08:34:00	1.69	13047	United Kingdom	sale	

	TotalAmount	MonthWeek	MonthofYear
0	15.30	W1	Dec
1	20.34	W1	Dec
2	22.00	W1	Dec
3	20.34	W1	Dec
4	20.34	W1	Dec
5	15.30	W1	Dec
6	25.50	W1	Dec
7	11.10	W1	Dec
8	11.10	W1	Dec
9	54.08	W1	Dec

InvoiceNo	int64
StockCode	int64
Description	object
Quantity	int64
InvoiceDate	object
UnitPrice	float64
CustomerID	int64
Country	object
TransactionType	object
TotalAmount	float64
MonthWeek	object
MonthofYear	object
dtype:	object

```

InvoiceNo          int64
StockCode          int64
Description         object
Quantity           int64
InvoiceDate        datetime64[ns]
UnitPrice          float64
CustomerID         int64
Country            object
TransactionType     object
TotalAmount        float64
MonthWeek          object
MonthofYear        object
dtype: object

```

```

      InvoiceNo      StockCode      Quantity \
count  399654.000000  399654.000000  399654.000000
mean    560559.445716    30774.912587    13.569453
min      536365.000000    10002.000000     1.000000
25%      549148.000000    22029.000000     2.000000
50%      561857.000000    22666.000000     6.000000
75%      572035.000000    23268.000000    12.000000
max      581587.000000    90214.000000   80995.000000
std       13087.440133    20616.684717    250.778916

```

```

      InvoiceDate      UnitPrice      CustomerID \
count              399654  399654.000000  399654.000000
mean  2011-07-10 12:46:01.383446784      2.907621  15288.681555
min      2010-12-01 08:26:00      0.000000  12346.000000
25%      2011-04-06 15:31:00      1.250000  13959.000000
50%      2011-07-29 16:24:00      1.950000  15152.000000
75%      2011-10-20 12:03:00      3.750000  16791.000000
max      2011-12-09 12:50:00     649.500000  18287.000000
std              NaN      4.452040   1710.807345

```

```

      TotalAmount
count  399654.000000
mean      23.041940
min        0.000000
25%        4.950000
50%       11.900000
75%       19.800000
max    168469.600000
std      425.412835

```

3888

1.1.1 1. Plotting the Proportion of Transaction Types Captured as either Sale and Canceled

	TransactionType	TotalAmount
0	canceled	471742.31
1	sale	8737061.14

1.1. Visualizing the above result using a pie chart

Retail Sales Transaction Type by Proportion(%)



1.2. Insights: From the above pie chart, it is clear **cancelled** transactions make up the smallest percentage of all transaction on the online retail platform. Occupying just 5.12% of all transactions, it accumulated a total amount of 471,742.31. Transactions that were captured as **Sale** make up the largest percentage of all transaction. Occupying 94.9% and accumulated a total amount of 8,737,061.14.

1.1.2 2. Monthly and Yearly Trend in Transactions With Respect to Transaction Types ('sale' or 'cancelled')

We will visualize the trend in sales for each month for the recorded period in our dataset. Firstly, we will look at transactions that are categorized as **sale** and look at the trend within this category throughout the recorded period. Then we will look at transactions that are categorized as **cancelled**. For this trend analysis, we will utilize a area plot.

2.1. Extract month and year into separate columns

2.2. Filter for transactions of type 'sale'

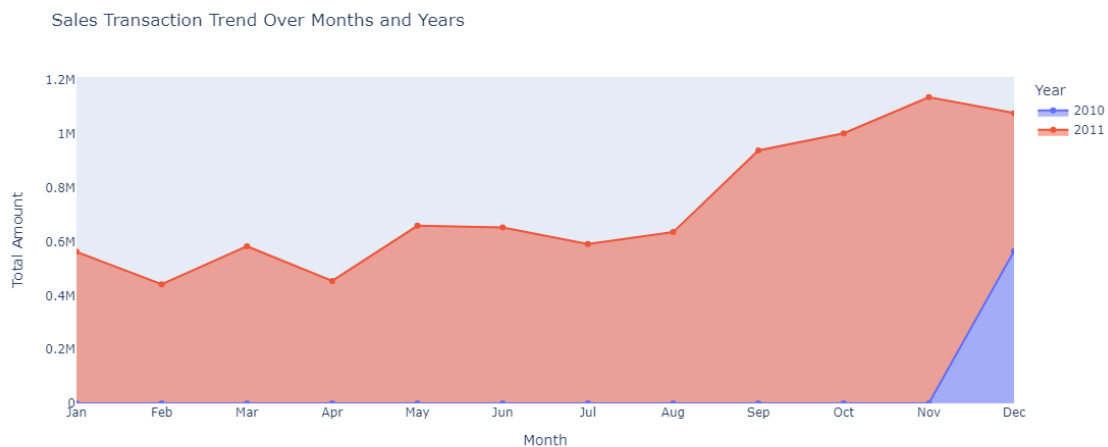
2.3. Create a new DataFrame with the desired aggregation

2.4. Order the DataFrame by year and month

	Year	Month	TotalAmount
0	2010	Dec	565145.06
5	2011	Jan	562657.71
4	2011	Feb	442291.09

8	2011	Mar	583121.35
1	2011	Apr	454435.88
9	2011	May	659220.69
7	2011	Jun	653263.67
6	2011	Jul	591600.04
2	2011	Aug	635504.38
12	2011	Sep	938736.38
11	2011	Oct	1002324.06
10	2011	Nov	1136532.75
3	2011	Dec	512228.08

2.5. Visualizing the sales transactions trend of the above result set using an area plot



2.6. Insights: From the trend in transactions categorized as **sale**, only the month of December in the year 2010 was sales recorded. For that month an amount of **565,145K** was recorded. Although no data on sales for the other months of that year 2010 were not available of this analysis. Comparing it to the same month in the year 2011 showed a slightly better performance in sales in 2010 than in 2011. For sales in 2011, we observed a decline in the months of **February** and **April** with recorded sales figures of **442,291K** and **454,436K** respectively. A rise in sales occurred from the month of **August** through to **November**, with **November** recording the highest sales in the year with a total sales amount of **1.137 million**.

2.7. Filter for Transactions of type ‘canceled’

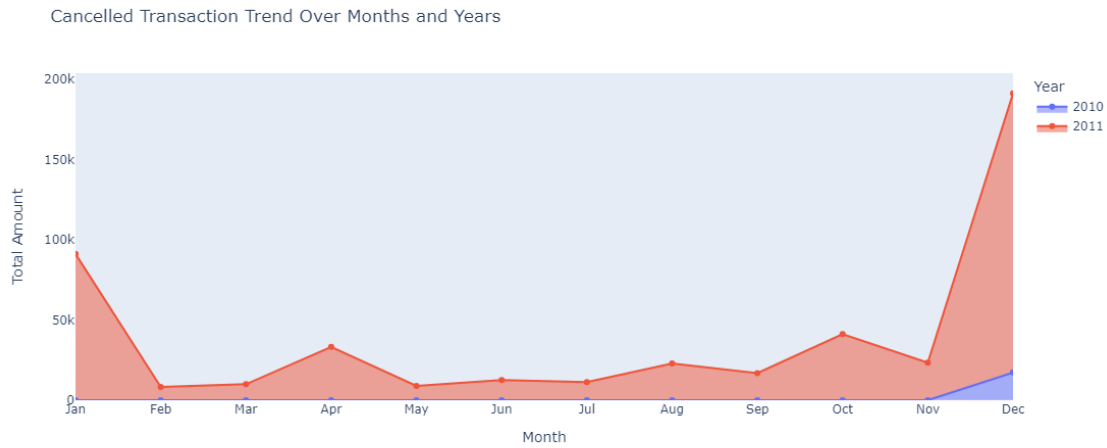
2.8. Create a new data frame with desired aggregation

2.9. Ordering the data frame by Year and Month

	Year	Month	TotalAmount
0	2010	Dec	17297.24

5	2011	Jan	91252.57
4	2011	Feb	8275.42
8	2011	Mar	10005.80
1	2011	Apr	33313.11
9	2011	May	8928.25
7	2011	Jun	12635.71
6	2011	Jul	11389.06
2	2011	Aug	22909.25
12	2011	Sep	16943.89
11	2011	Oct	41284.07
10	2011	Nov	23431.88
3	2011	Dec	174076.06

2.10. Visualizing the cancelled transactions trend of the above data frame using an area plot



2.11. Insights: From the trend in transactions categorized as **canceled**, only the month of December in the year 2010 was any cancelled transactions recorded. For that month an amount of **17,297K** was recorded. No data on cancelled transactions for the other months of that year (2010) were not available of this analysis. Comparing it to the same month in the year 2011 showed a huge increase in the total amount recorded **174,076K** for cancelled transactions. For 2011, the retail store observed its lowest cancelled transactions in the months of **February** and **May** with recorded figures of **8,275K** and **8,928K** respectively. The highest amount of total cancelled transactions occurred in the **December**, recording the highest total amount of **174,076K**.

1.1.3 3. Countries with the most sales and canceled Transactions

In the dataset from the online retail store, the countries from which customers transacted were captured. This information is going to be used to derive insights into the total amounts realised from each country with respect to the type of transaction category. We will look at the top 10 countries

with the highest sales transacted and the bottom 10 countries with the least sales transacted on the online retail store. Then the top 10 countries with the highest cancelled transactions and the bottom 10 countries with the least cancelled transactions. Bar charts will be used to visualize and communicate the results of this analysis.

3.1. Filtering for transactions of type sale

3.2. Grouping by 'Country' and aggregating by the number of Transactions and Total Amount

3.3. Sorting the countries by number of transactions in descending order to find the top countries

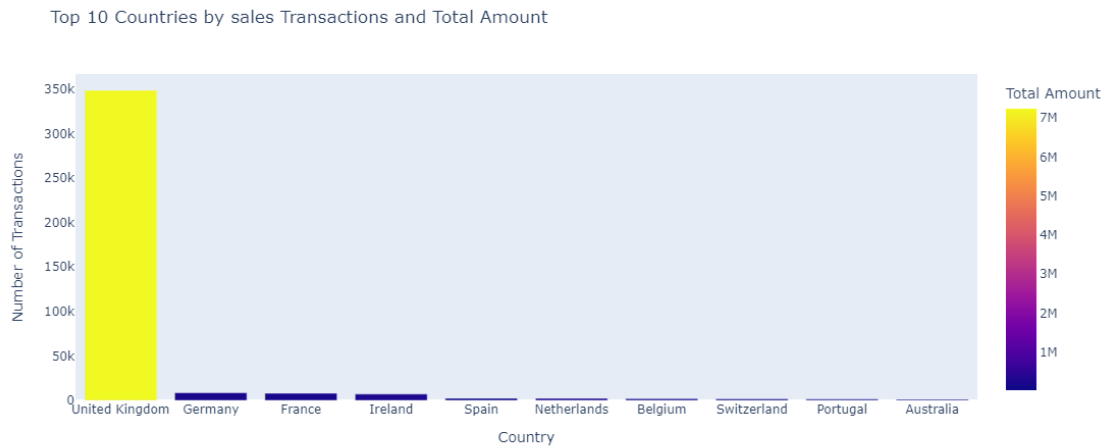
	Country	TransactionCount	TotalAmount
35	United Kingdom	348860	7242707.22
13	Germany	8644	205381.15
12	France	8019	183801.94
16	Ireland	7128	257013.12
30	Spain	2418	55706.56
23	Netherlands	2326	283889.34
3	Belgium	1935	36927.34
32	Switzerland	1811	52441.95
26	Portugal	1416	26643.90
0	Australia	1183	138103.81
24	Norway	1049	32454.64
6	Channel Islands	743	19997.54
18	Italy	741	15820.24
11	Finland	647	18344.88
7	Cyprus	600	13158.76
31	Sweden	427	36828.83
1	Austria	384	8742.68
9	Denmark	367	18211.34
25	Poland	325	6974.65
19	Japan	321	37416.37
17	Israel	244	7205.84
36	Unspecified	241	2660.77
28	Singapore	215	9120.39
15	Iceland	182	4310.00
33	USA	179	3580.39
5	Canada	150	3115.44
14	Greece	142	4425.52
22	Malta	109	2070.59
34	United Arab Emirates	67	1864.78
10	EU	57	1159.25
29	South Africa	57	1002.31
20	Lebanon	45	1693.88
21	Lithuania	35	1661.06
4	Brazil	32	1143.60

8	Czech Republic	24	786.74
2	Bahrain	17	548.40
27	Saudi Arabia	9	145.92

3.4. Top 10 Countries for Sales Transactions

	Country	TransactionCount	TotalAmount
35	United Kingdom	348860	7242707.22
13	Germany	8644	205381.15
12	France	8019	183801.94
16	Ireland	7128	257013.12
30	Spain	2418	55706.56
23	Netherlands	2326	283889.34
3	Belgium	1935	36927.34
32	Switzerland	1811	52441.95
26	Portugal	1416	26643.90
0	Australia	1183	138103.81

3.5. Visualizing the above results Due to the number of countries in the result set, the top 10 countries with respect to number of transactions and total `sales` amount, will be considered in the visualization first and then the bottom 10 countries.



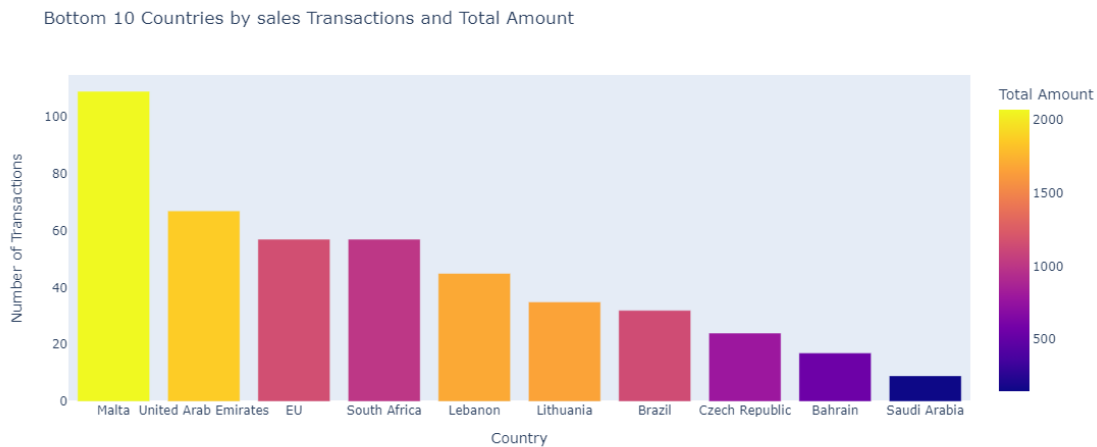
3.6. Insights: For the most `sale` transactions recorded on the online retail store with respect to the countries of residence of customers, the **United Kingdom** recorded the highest number of `sales` transactions. The total figure recorded was approximately **348,800** transactions of the category `sale`. The **United Kingdom** also recorded the highest total amount with respect to sales with a total amount of **7.2 million**. By far this was the highest recorded sales amount for all transactions on the online retail store platform with respect to countries. In the chart above, we selected only the top 10 countries, with Australia having the lowest number of transactions with respect to sales (**1183**) and the lowest total amount **138,103.81**. The top 10 countries in

the result set were; **United Kingdom, Germany, France, Ireland, Spain, Netherlands, Belgium, Switzerland, Portugal and Australia** in that order.

3.7. Bottom 10 Countries for Sales Transactions

	Country	TransactionCount	TotalAmount
22	Malta	109	2070.59
34	United Arab Emirates	67	1864.78
10	EU	57	1159.25
29	South Africa	57	1002.31
20	Lebanon	45	1693.88
21	Lithuania	35	1661.06
4	Brazil	32	1143.60
8	Czech Republic	24	786.74
2	Bahrain	17	548.40
27	Saudi Arabia	9	145.92

3.8. Visualizing the above results



3.9. Insights: For the bottom 10 countries with the least **sale** transactions recorded on the online retail store platform, the **Saudia Arabia** recorded the least or lowest number of **sales** transactions. With a total number of transactions of just **9** transactions of the category **sale**. **Saudi Arabia** also recorded the lowest total amount with respect sales with a total amount of just **145.92**. By far this was the lowest recorded sales amount for all transactions on the online retail store platform with respect to countries. In the bar chart above, we selected only the bottom 10 countries, with Malta having the highest number of transcsations with respect to sales, **109** and a total amount of **2,070.59** among the group. The bottom 10 countries in the result set were; **Malta, United Arab Emirates, EU, South Africa, Lebanon, Lithuania, Brazil, Czech Republic, Bahrain and Saudi Arabia** in that order. It must be noted that among the bottom 10 group, we had The E.U captured as a country. This might be due to customers preferring to

use their geographical region instead of the country of residence. This will have to be looked at in the design of the customer data capture form on the platform to ensure data conformity.

3.10. Filtering for transactions of type canceled

3.11. Grouping by ‘Country’ and aggregating by number Transactions and Total Amount

3.12. Sorting the DataFrame by number of transactions(TransactionCount) in descending order to find the top countries

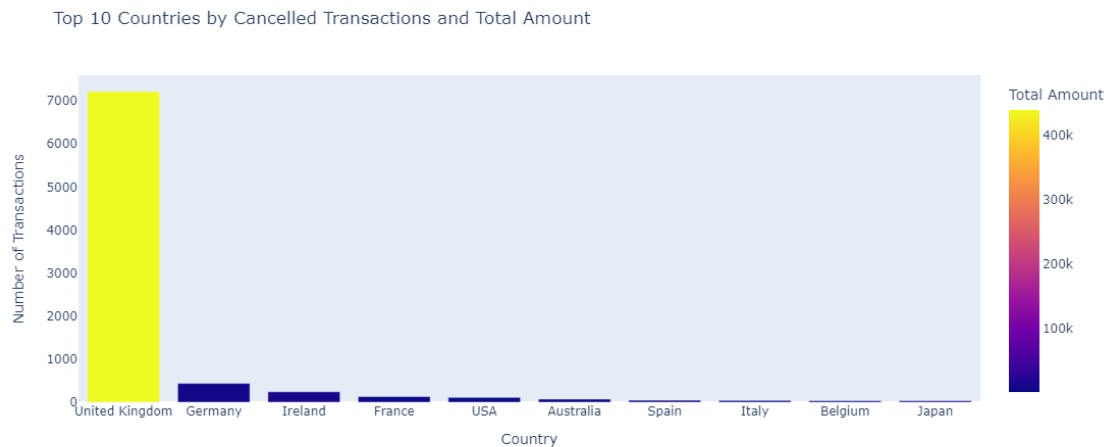
	Country	TransactionCount	TotalAmount
25	United Kingdom	7217	440301.34
10	Germany	437	4761.49
11	Ireland	242	10280.65
9	France	133	2416.40
24	USA	112	1849.47
0	Australia	73	1181.31
21	Spain	45	3959.91
13	Italy	42	543.90
2	Belgium	36	264.38
14	Japan	34	1996.58
23	Switzerland	33	582.55
15	Malta	14	90.12
17	Norway	11	161.68
19	Portugal	11	143.64
18	Poland	11	121.51
22	Sweden	9	1662.42
3	Channel Islands	9	71.15
6	Denmark	8	169.20
4	Cyprus	7	323.40
8	Finland	6	41.34
5	Czech Republic	4	115.02
16	Netherlands	4	409.80
1	Austria	3	44.36
12	Israel	2	227.44
7	EU	1	8.50
20	Saudi Arabia	1	14.75

3.13. Top 10 Countries for Canceled Transactions

	Country	TransactionCount	TotalAmount
25	United Kingdom	7217	440301.34
10	Germany	437	4761.49
11	Ireland	242	10280.65
9	France	133	2416.40
24	USA	112	1849.47
0	Australia	73	1181.31
21	Spain	45	3959.91

13	Italy	42	543.90
2	Belgium	36	264.38
14	Japan	34	1996.58

3.14. Visualizing the above result using a bar chart Due to the number of countries in the result set, the top 10 countries with respect to number of transactions and total canceled transaction amount, will be considered in the visualization first and then the bottom 10 countries.



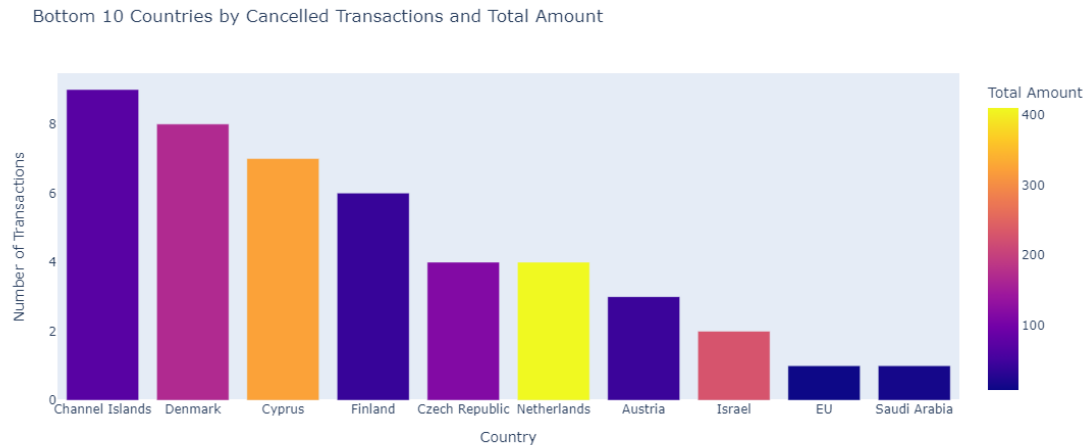
3.15. Insights: From the visualization above, among the top 10 countries with the most canceled transactions recorded on the online retail store platform, the **United Kingdom** recorded the highest number of canceled transactions. With a total number of canceled transactions at **7217** transactions of the category canceled. The **United Kingdom** also recorded the highest total amount with respect cancelled transactions, with a total amount of **440,301.34**. By far this is the highest recorded canceled transaction amount for all transactions on the online retail store platform with respect to countries. Of the top 10 countries with the highest among of canceled transactions, **Japan** recorded the least highest, with a total number of canceled transactions of **34** and a total amount of **1,996.58**. The top 10 countries in the result set are; **United Kindgom, Germany, Ireland, France, USA, Australia, Spain, Italy, Belgium and Japan** in that order.

3.16. Bottom 10 Countries for Canceled Transactions

	Country	TransactionCount	TotalAmount
3	Channel Islands	9	71.15
6	Denmark	8	169.20
4	Cyprus	7	323.40
8	Finland	6	41.34
5	Czech Republic	4	115.02
16	Netherlands	4	409.80
1	Austria	3	44.36

12	Israel	2	227.44
7	EU	1	8.50
20	Saudi Arabia	1	14.75

3.17. Visualizing the above result using a bar chart



3.18. Insights: For the bottom 10 countries with the least canceled transactions recorded on the online retail store platform, the **Saudia Arabia** and the **EU** recorded the least or lowest number of canceled transactions. With a total number of transactions of just **1** transaction each. Although **Saudi Arabia** and the **EU** recorded the least number of canceled transactions, in terms of total amount, **Saudi Arabia** recorded an amount of **14.75** which is higher than that of the **EU** which stands at just **8.50**. By far this was the lowest recorded sales amount for all cancelled transactions on the online retail store platform with respect to countries. In the bar chart above, we selected only the bottom 10 countries, with the **Channel Islands** having the highest number of cancelled transactions with **9** and a total amount of **71.15** among the group. The bottom 10 countries in the result set were; **Channel Islands, Denmark, Cyprus, Finland, Czech Republic, Netherlands, Austria, Israel, EU, and Saudi Arabia** in that order. It must be noted that among the bottom 10 group, we had The E.U also captured as a country. This might be due to customers preferring to use their geographical region instead of the country of residence.

1.1.4 4. During which time of day were the transactions made

For this analysis, I will look at both transaction types **sale** and **canceled**, and delve into the dataset to derive insights from transactions made during the times of day, ie. Morning, Afternoon and Evening. We will start a top level, looking at this analysis with respect to months and drill down with filters to look at it with respect to weeks for each year (2010 and 2011).

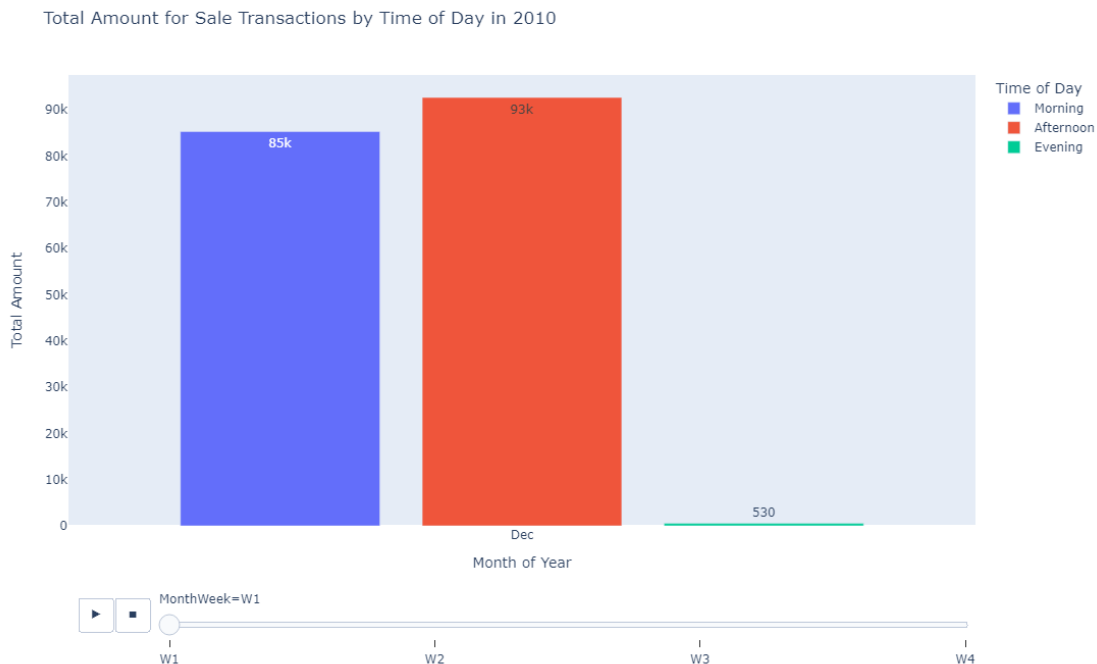
4.1. Filter out all transcatons that are categorized as sale for the year 2010

4.2. Create new column for the time of day and categorize into morning, afternoon or evening

4.3. Group the sales transaction data by week, month and year for 2010

	Year	Month	MonthWeek	TimeOfDay	TotalAmount
0	2010	Dec	W1	Morning	85285.76
1	2010	Dec	W1	Afternoon	92628.70
2	2010	Dec	W1	Evening	529.55
3	2010	Dec	W2	Morning	97282.40
4	2010	Dec	W2	Afternoon	109003.40
5	2010	Dec	W2	Evening	1022.13
6	2010	Dec	W3	Morning	80092.46
7	2010	Dec	W3	Afternoon	71022.05
8	2010	Dec	W3	Evening	2373.26
9	2010	Dec	W4	Morning	8286.45
10	2010	Dec	W4	Afternoon	17618.90

4.4. Visualizing the above results for transactions categorized as sale in 2010



4.5. **Insights:** For the year 2010, transactions on the online retail store were captured only for the month of **December**. In the four weeks of that month, the following insights were derived with respect to the total amount recorded for the transaction type **sale**.

- In week one (W1), more sales transactions were made in the afternoon with a recorded total amount of **92,628.70** as compared to the other times of the day in that week. Evening time

recorded the least or lowset amount of sales transactions with a recorded total amount of just **529.55**.

- In week two (W2), more sales transactions were made in the afternoon with a recorded total amount of **109,003.40** in that week. The morning time recorded the next highest with a total amount of **97,282.40**, with the evening time recording the least sales transactions with a recorded total amount of **1,022.13**.
- In week three (W3), the mornings recorded the most sales transactions with total amount of **80,092.46** in that week. The afternoon time recorded the next highest with a total amount of **71,022.05**, with the evening time recording the least sales transactions with a recorded total amount of **2,373.26**.
- In last week of the month of **December**, week four (W4), less sales transactions were made in all the times of day of the month as compared to the previous weeks in the same month. The afternoon time recorded highest total amount of **17,618.90** in that week. The morning time recorded the next highest with a total amount of **8,286.45**, with the evening time recording the least sales transactions with a recorded total amount of **2,373.26**.
- It must be noted that the total amount recorded for transactions of type **sale** for the evening times in both week three and four was the same, **2,373.26**.

4.6. Filter out all transcatons that are categorized as sale for the year 2011

4.7. Create new column for the time of day and categorize into morning, afternoon or evening

	InvoiceNo	StockCode	Description	Quantity	\
26230	539993	22386	JUMBO BAG PINK POLKADOT	10	
26231	539993	21499	BLUE POLKADOT WRAP	25	
26232	539993	21498	RED RETROSPOT WRAP	25	
26233	539993	22379	RECYCLING BAG RETROSPOT	5	
26234	539993	20718	RED RETROSPOT SHOPPER BAG	10	
26235	539993	85099	JUMBO BAG RED RETROSPOT	10	
26236	539993	20682	RED RETROSPOT CHILDRENS UMBRELLA	6	
26237	539993	22961	JAM MAKING SET PRINTED	12	
26238	539993	22667	RECIPE BOX RETROSPOT	6	
26239	539993	22898	CHILDRENS APRON APPLES DESIGN	8	

	InvoiceDate	UnitPrice	CustomerID	Country	\
26230	2011-01-04 10:00:00	1.95	13313	United Kingdom	
26231	2011-01-04 10:00:00	0.42	13313	United Kingdom	
26232	2011-01-04 10:00:00	0.42	13313	United Kingdom	
26233	2011-01-04 10:00:00	2.10	13313	United Kingdom	
26234	2011-01-04 10:00:00	1.25	13313	United Kingdom	
26235	2011-01-04 10:00:00	1.95	13313	United Kingdom	
26236	2011-01-04 10:00:00	3.25	13313	United Kingdom	
26237	2011-01-04 10:00:00	1.45	13313	United Kingdom	
26238	2011-01-04 10:00:00	2.95	13313	United Kingdom	
26239	2011-01-04 10:00:00	1.95	13313	United Kingdom	

	TransactionType	TotalAmount	MonthWeek	MonthofYear	Month	Year	TimeOfDay
26230	sale	19.5	W1	Jan	Jan	2011	Morning
26231	sale	10.5	W1	Jan	Jan	2011	Morning
26232	sale	10.5	W1	Jan	Jan	2011	Morning
26233	sale	10.5	W1	Jan	Jan	2011	Morning
26234	sale	12.5	W1	Jan	Jan	2011	Morning
26235	sale	19.5	W1	Jan	Jan	2011	Morning
26236	sale	19.5	W1	Jan	Jan	2011	Morning
26237	sale	17.4	W1	Jan	Jan	2011	Morning
26238	sale	17.7	W1	Jan	Jan	2011	Morning
26239	sale	15.6	W1	Jan	Jan	2011	Morning

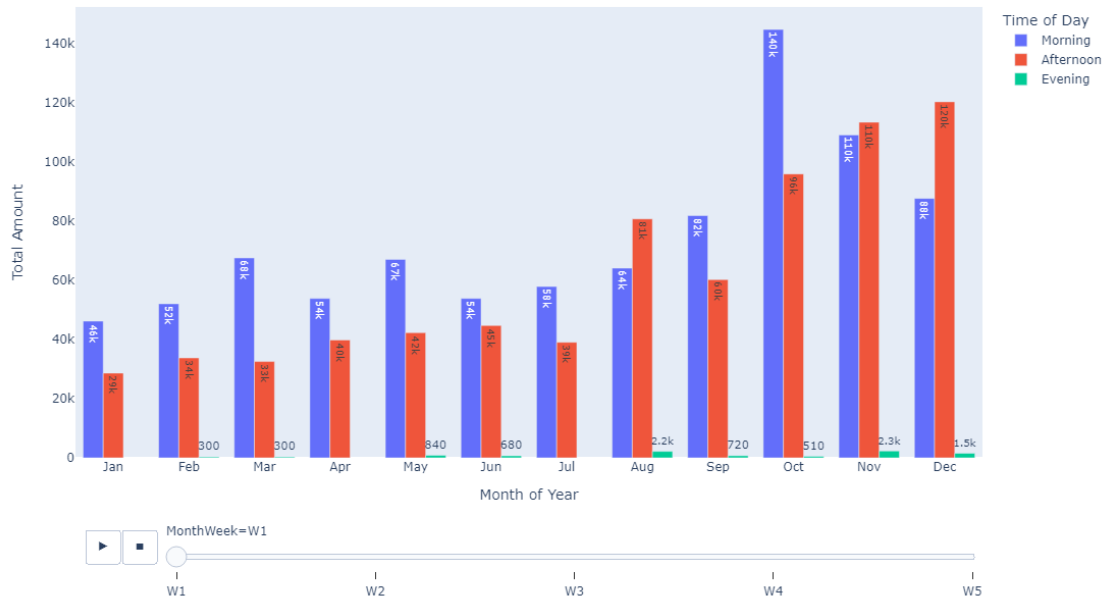
4.8. Group the sales transaction data by week, month and year for 2011

	Year	Month	MonthWeek	TimeOfDay	TotalAmount
0	2011	Jan	W1	Morning	46173.53
1	2011	Jan	W2	Morning	101976.37
2	2011	Jan	W3	Morning	136559.55
3	2011	Jan	W4	Morning	59907.43
4	2011	Jan	W5	Morning	25458.44
..
153	2011	Dec	W2	Morning	242274.01
154	2011	Dec	W1	Afternoon	120246.63
155	2011	Dec	W2	Afternoon	59483.99
156	2011	Dec	W1	Evening	1531.81
157	2011	Dec	W2	Evening	1078.96

[158 rows x 5 columns]

4.9. Visualizing the above results for transactions categorized as sale in 2011

Total Amount for Sale Transactions by Time of Day in 2011



4.10. Insights: For the year 2011, transactions on the online retail store were captured for the months of **January** through to **December**. In the five weeks of each month of the year, the following insights were derived with respect to the total amount recorded for the transaction type **sale**.

- In week one (W1), **January** didn't record any transactions in the evening. Across all the months in the year, the evening times recorded the least total amount for transactions made on the online retail store. The months of **February** and **March** recorded the least total_amounts for sales transactions for the evenings in the year, with a total amount of **300** each. Comparatively, more sales transactions were made in the mornings across all the months of the year with **October** recording the highest total amount of **144,624.77** for all morning sales transactions. **August**, **November** and **December** were the only months in which a higher total amount with respect to sales transactions were made in the afternoon, as compared to the other times of the day in that week. The recorded total amounts for sales transactions for the respective months were **80,678.39** for **August**, **113,325.11** for **November** and **120,246.63** for **December**.
- In week two (W2), **January**, **August** and **October** didn't record any total amount for sales transactions in the evening. Across all the months in the year, the evening times recorded the least total amount for transactions made on the online retail store. The months of **February** and **March** recorded the least total_amounts for sales transactions for the evenings in the year, with a total amounts of **220** and **170** each respectively. Across all the months in the year, the mornings recorded the highest total amounts for sales transactions with **December** recording the highest total amount of **242,274.01**. **June**, **July** and **August** were the only months in which a higher total amount with respect to sales transactions were made in the afternoon, as compared to the other times of the day in that week. The recorded total amounts for sales

transactions for these respective months were **119,584.12** for June, **54,302.36** for July and **82,972.06** for August.

- In week three (W3), no sales transactions were recorded for the month of **December** for each time of the day (morning, afternoon and evening). Again across all the months in the year for this week, the evening times recorded the least total amount for sales transactions made on the online retail store. The months of **January**, **February** and **September** recorded the least total_ amounts for sales transactions for the evenings in the year, with total amounts of **222.24**, **359.83** and **176.94** respectively. The mornings recorded the highest total amounts for sales transactions across all the months of the year, with **September** recording the highest total amount of **174,271.02**. **April** was the only month in which a higher total amount with respect to sales transactions was made in the afternoon, as compared to the other times of the day in that week. With a recorded total amount of **81,784.52**.
- In week four (W4), no sales transactions were recorded for the month of **December** for each time of the day (morning, afternoon and evening). Again across all the months in the year for this week, the evening times recorded the least total amounts for sales transactions made on the online retail store. The months of **January**, **March**, **July**, **August**, **September** and **November** recorded the least total_ amounts for sales transactions for the evenings in the year, with total amounts of **153.15**, **349.23**, **533.57**, **353.45**, **715.81** and **785.23** respectively less than a **1000**. The mornings recorded the highest total amounts for sales transactions across all the months of the year, with **October** recording the highest total amount of **127,350.80**. **March**, **September** and **November** were the only months in which a higher total amount with respect to sales transactions were recorded in the afternoon, as compared to the other times of the day in that week. With recorded total amounts of **66,719.99**, **108,349.98**, **122,708.20** respectively.
- In week five (W5), no sales transactions were recorded for the month of **December** for each time of the day (morning, afternoon and evening). Again across all the months in the year for this week, the evening times recorded the least total amounts for sales transactions made on the online retail store. **January**, **February**, **May**, **August**, **October** and **November** recorded no sales transactions for the evening time. The months of **March**, **April**, **July** and **September** recorded the least total_ amounts for sales transactions for the evenings in the year, with total amounts of **984.90**, **94.45**, **394.85** and **243.20** respectively less than a **1000**. The mornings recorded some of the highest total amounts for sales transactions in major of the months in year, with **September** recording the highest total amount of **67,027.75**. **May**, **July** and **November** were the only months in which higher total amounts with respect to sales transactions were recorded in the afternoon, as compared to the other times of the day in that week. With recorded total amounts of **13,007.18**, **73,531.56**, **75,534.87** respectively.

4.11. Filter out all transctions that are categorized as canceled for the year 2010

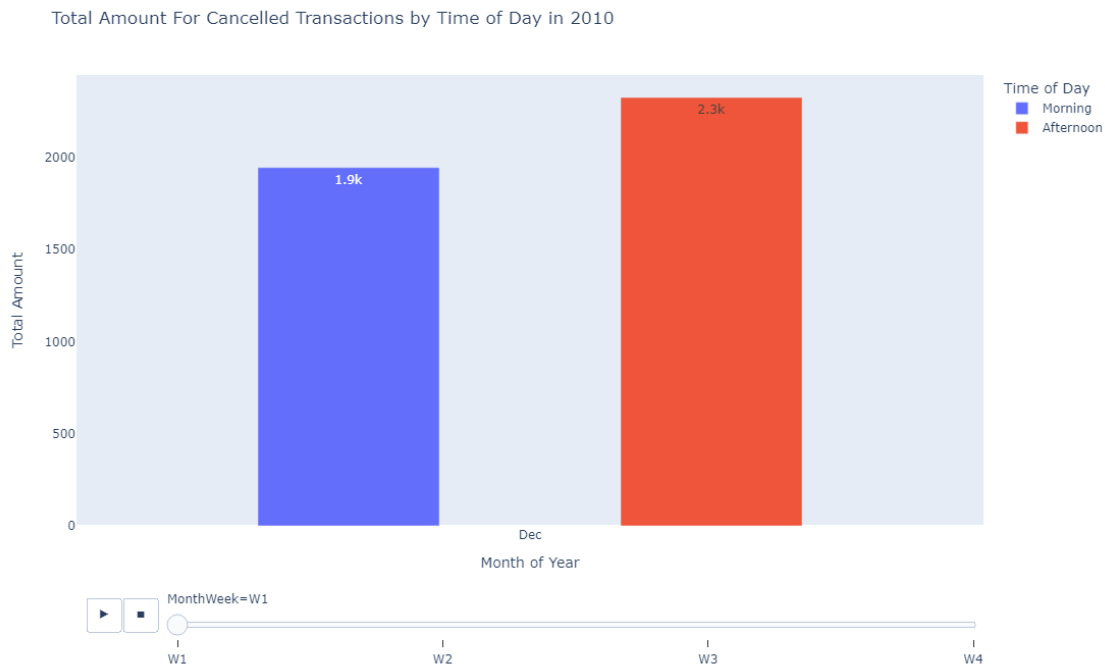
4.12. Create new column for the time of day and categorize into morning, afternoon or evening

4.13. Group the canceled transactions data by week, month and year for 2010

	Year	Month	MonthWeek	TimeOfDay	TotalAmount
0	2010	Dec	W1	Morning	1944.84

1	2010	Dec	W1	Afternoon	2325.10
2	2010	Dec	W2	Morning	588.52
3	2010	Dec	W2	Afternoon	3671.20
4	2010	Dec	W3	Morning	3923.41
5	2010	Dec	W3	Afternoon	4326.92
6	2010	Dec	W4	Morning	394.95
7	2010	Dec	W4	Afternoon	122.30

4.14. Visualizing the above results for transactions categorized as canceled in 2010



4.15. Insights: For the year 2010, transactions of type **canceled** on the online retail store were captured only for the month of **December**. In the four weeks of that month, the following insights were derived with respect to the total amount recorded for the transaction type **canceled**.

- In week one (W1), more cancelled transactions were recorded in the afternoon with a total amount of **2,325.10** as compared to the morning which recorded a total amount of **1,944.84** in that week. There were no recorded cancelled transactions for evening time in the month.
- In week two (W2), more cancelled transactions were recorded in the afternoon with a total amount of **3,671.20** in that week. The morning time recorded a total amount of just **588.52**, with no cancelled transactions recorded in the evening time.
- Week three (W3) recorded the most cancelled transactions for both morning and afternoon, with total amounts of **3,923.41** and **4,326.92** respectively in that week. The afternoon time recorded the highest total amount of the two, with the evening time recording no cancelled transactions.

- The last week of the month, week four (W4), recorded the least cancelled transactions in all the times of day of the month as compared to the previous weeks in the same month. The afternoon time recorded the least total amount of **122.30**. The morning time recorded the highest with a total amount of **394.95**, again no cancelled transaction for the evening time was recorded.

4.16. Filter out all transactions that are categorized as canceled for the year 2011

4.17. Create new column for the time of day and categorize into morning, afternoon or evening

	InvoiceNo	StockCode	Description	Quantity	\
26298	540006	21306	SET/4 DAISY MIRROR MAGNETS	1	
26299	540006	84352	SILVER CHRISTMAS TREE BAUBLE STAND	1	
26300	540006	22423	REGENCY CAKESTAND 3 TIER	1	
26301	540007	21055	TOOL BOX SOFT TOY	6	
26302	540007	22274	FELTCRAFT DOLL EMILY	6	

	InvoiceDate	UnitPrice	CustomerID	Country	\
26298	2011-01-04 10:48:00	2.10	14606	United Kingdom	
26299	2011-01-04 10:48:00	16.95	14606	United Kingdom	
26300	2011-01-04 10:48:00	12.75	14606	United Kingdom	
26301	2011-01-04 11:08:00	8.95	15379	United Kingdom	
26302	2011-01-04 11:08:00	2.95	15379	United Kingdom	

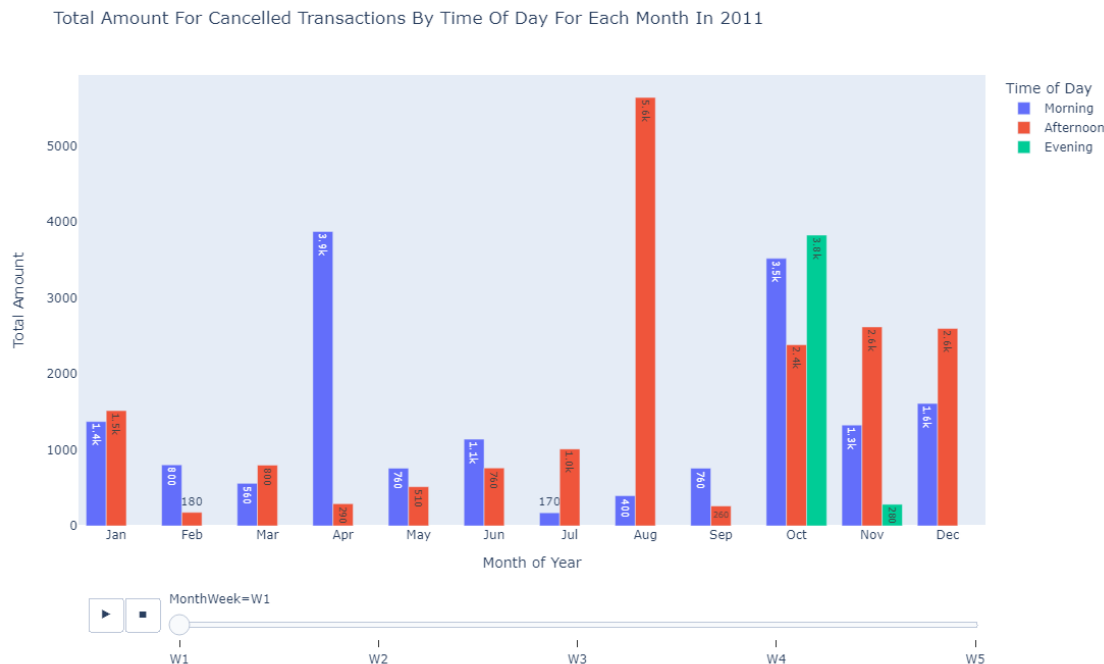
	TransactionType	TotalAmount	MonthWeek	MonthofYear	Month	Year	TimeOfDay
26298	canceled	2.10	W1	Jan	Jan	2011	Morning
26299	canceled	16.95	W1	Jan	Jan	2011	Morning
26300	canceled	12.75	W1	Jan	Jan	2011	Morning
26301	canceled	53.70	W1	Jan	Jan	2011	Morning
26302	canceled	17.70	W1	Jan	Jan	2011	Morning

4.18. Group the canceled transactions data by week, month and year for 2011

	Year	Month	MonthWeek	TimeOfDay	TotalAmount
0	2011	Jan	W1	Morning	1373.13
1	2011	Jan	W2	Morning	2992.18
2	2011	Jan	W3	Morning	77597.01
3	2011	Jan	W4	Morning	1183.81
4	2011	Jan	W5	Morning	2018.15
..
121	2011	Dec	W1	Morning	1609.73
122	2011	Dec	W2	Morning	169418.71
123	2011	Dec	W1	Afternoon	2596.74
124	2011	Dec	W2	Afternoon	380.73
125	2011	Dec	W2	Evening	70.15

[126 rows x 5 columns]

4.19. Visualizing the above results for canceled transactions in 2011



4.20. Insights: For the year 2011, transactions on the online retail store were captured for the months of January through to December. In the five weeks, the following insights were derived with respect to the total amount recorded for the transactions of type canceled.

- **Week One (W1):**

1. For canceled transactions in the morning, April recorded the highest total amount **3,874.67**, with the month of July recording the least total amount of just **171.67**.
2. For the afternoon, the months of January, March, July, August, November and December recorded higher total amounts for cancelled transactions. The month of **August** recorded the highest total amount for **canceled** transactions among the listed months with a recorded total amount of **5,637.25**. **February** had the least total amount for **canceled** transactions with a figure of just **177.37**.
3. In the evening, the months of January, February, March, April, May, June, July, August, September and December didn't record any cancelled transactions. The only months for which any cancelled transactions in the evening were recorded, were **October** and **November**. In these two months, **October** recorded the highest total amount of **3,826.75**. This figure was the highest in the month, higher than the total amounts for the morning and afternoon.
4. For Ten months out of the 12, no canceled transactions were recorded for the evening. These months were; January, February, March, April, May, June, July, August, September, and December.

- **Week Two (W2):**

1. For canceled transactions in the morning, the months of October and December recorded

the highest total amounts for all canceled transactions with figures of **13,629.41** and **169,418.71** respectively. The month of December recorded the highest total amount of the week. The months of March, August and September recorded the least total amount of canceled transactions.

2. For the afternoon, the months of October, and November recorded higher total amounts for canceled transactions. The month of **October** recorded the highest total amount for canceled transactions of the two months with a recorded amount of **3,369.90**. May had the least total amount for **canceled** transactions with a figure of just **315.28**.
3. In the evening, only the months of May, June, and December recorded any cancelled transactions. Of these three months, **June** recorded the highest total amount of **226.26**, followed by **May 223.85** with **December** recording the least total amount of just **70.15**.
4. Nine(9) months out of the 12, recorded no canceled transactions for the evening. These months were; January, February, March, April, July, August, September, October, and November.

- **Week Three (W3):**

1. For canceled transactions in the morning, January recorded the highest total amount **77,597.01**, with the month of February recording the least total amount of just **311.70**. For the month of **December**, no canceled transactions were recorded for all times of the day in the week.
2. For the afternoon, the month of April recorded the highest total amount for cancelled transactions with a figure of **23,956.02**. The month of **February** recorded the least total amount for canceled transactions of all the months with a recorded amount of just **441.64**.
3. For the evening, the only months for which any cancelled transactions were recorded, were **May, October** and **November**. In these three months, **November** recorded the highest total amount of just **165**, with **May** recording the least amount of **61.85**. Of all the times of day for this week, the evening recorded the least total amounts for all canceled transactions.
4. For eight months out of the eleven available months, no canceled transactions were recorded for the evening. These months were; January, February, March, April, June, July, August, and September. The month of December recorded no canceled transactions for each time of day for the week.

- **Week Four (W4):**

1. For the morning, September and October recorded the highest total amounts with figures of **7,387.56** and **6,132.71** respectively. With the month of September recording the highest total amount in the week. The month of **July** recorded the least total amount in the week with just **185.83**.
2. For the afternoon, the month of February recorded highest total amount for cancelled transactions with recorded total amount of **2,728.80**. **April** had the least total amount for **canceled** transactions with a figure of just **396.73**.
3. For the evening, only the months of March, May, and October recorded any cancelled transactions. For these three months, **May** recorded the highest total amount of **100.18**. Of all the times of day for the week, the evening time recorded the least total amounts for canceled transactions.
4. For eight months out of the available 11, no canceled transactions were recorded for the evening. These months were; January, February, April, June, July, August, September and November. The month of December recorded no canceled transactions for any time of day for the week.

- **Week Five (W5):**

1. For canceled transactions in the morning, the month of September recorded the highest total amount **3,289.20**, with the month of October recording the least total amount of just **239.50**.
2. For the afternoon, the month of August recorded highest total amount for cancelled transactions with a figure of **8,174.76**. The month of February had the least total amount for canceled transactions with a figure of just **1.95** and this was the only recorded total amount for February for the week.
3. For the evening, the months of January, February, March, May, June, August, September, October, and November didn't record any cancelled transactions. The only months for which any cancelled transactions in the evening were recorded, were April and July. Of these two months, April recorded the highest total amount of **535.71**.
4. For nine(9) months out of the available 11 months, no canceled transactions were recorded for the evening. These months were; January, February, March, May, June, August, September, October and November. The month of December recorded no canceled transaction of all times of day for the week.

1.1.5 5. From which country did the online retail store have the most customers?

5.1. Count of customers that visited the online retail store for each country

5.2. Sorting the resulting dataframe by the number of customers in descending order

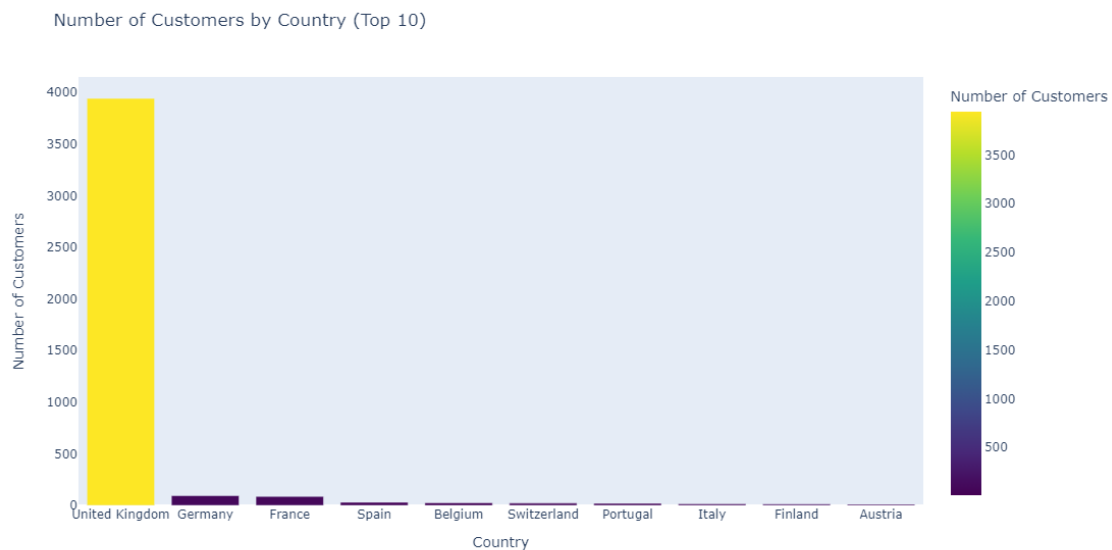
	Country	CustomerID
35	United Kingdom	3943
13	Germany	95
12	France	87
30	Spain	30
3	Belgium	25
32	Switzerland	21
26	Portugal	19
18	Italy	14
11	Finland	12
1	Austria	11
24	Norway	10
23	Netherlands	9
0	Australia	9
9	Denmark	9
6	Channel Islands	9
7	Cyprus	8
31	Sweden	8
19	Japan	8
25	Poland	6
33	USA	4
5	Canada	4
14	Greece	4
36	Unspecified	4
17	Israel	4

16	Ireland	3
22	Malta	2
34	United Arab Emirates	2
2	Bahrain	2
8	Czech Republic	1
21	Lithuania	1
20	Lebanon	1
27	Saudi Arabia	1
28	Singapore	1
29	South Africa	1
4	Brazil	1
15	Iceland	1
10	EU	1

5.3. Top 10 Countries with the Most Customers

	Country	CustomerID
35	United Kingdom	3943
13	Germany	95
12	France	87
30	Spain	30
3	Belgium	25
32	Switzerland	21
26	Portugal	19
18	Italy	14
11	Finland	12
1	Austria	11

5.4. Visualizing the above result For the resulting dataframe, the top 10 countries with the most customers will be visualized first and then the bottom 10 countries will be visualized.

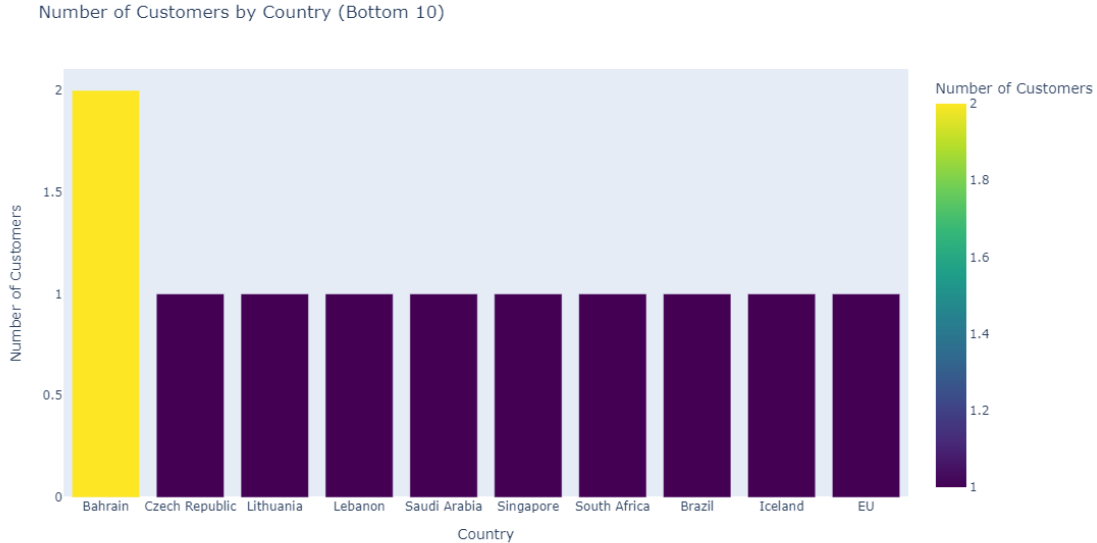


5.5. Insights: For the top 10 countries with the highest number of customers on the online retail store, the following insight was derived; 1. The United Kingdom had the highest number of customers with a total number of **3,943**. 2. Austria had the lowest number among the top 10 countries with a total number of **11**.

5.6. Bottom 10 Countries with Least Customers

	Country	CustomerID
2	Bahrain	2
8	Czech Republic	1
21	Lithuania	1
20	Lebanon	1
27	Saudi Arabia	1
28	Singapore	1
29	South Africa	1
4	Brazil	1
15	Iceland	1
10	EU	1

5.7. Visualizing the above result



5.8. Insights: For the bottom 10 countries with the least number of customers on the online retail store, the following insight was derived; 1. Bahrain had the highest number of customers among this group with a total number of just **2**. 2. The following countries, Czech Republic,

Lithuania, Lebanon, Saudi Arabia, Singapore, South Africa, Brazil, Iceland and the EU had the lowest number of customers of the bottom group with each country recording a total number of **1**.